

## NOTE

### UNIDIMENSIONAL DATA: A GRAPHICAL ANALYSIS

Mark F. BATELL

*The University of Michigan, Department of Mathematical Psychology, 530 Union Drive, Ann Arbor, MI 48109, USA*

Received 30 August 1980  
Revised 7 October 1980

The purpose of this article is three-fold: to demonstrate the interface between the theory of scaling and graph theory, to determine the class of unidimensional graphs, and to provide a method for analyzing data which consists of both dominance and indifference relations.

The most definitive work on the subject of data analysis is Coombs' (1976) *A Theory of Data* in which eight types of data are described. The eight data types arise from three defining dichotomies, one of which separates the data in terms of whether the obtained relations demonstrate dominance or indifference (termed proximity by Coombs). Let us assume, without loss of generality, that we have decided upon the method of paired comparisons for data collection. Under usual circumstances, when presented with two stimuli (or sets of stimuli, hereafter referred to simply as stimuli) the subject is asked to answer one of two questions. Either the subject must indicate which of the two stimuli dominates (is preferred to, is greater than, contains more of some aspect, etc.) the other, or the subject is asked whether or not the stimuli are proximal (similar, indistinguishable, within some indifference threshold of one another, etc.). There are numerous ways of handling data thus obtained (cf. Coombs, 1976 for a complete treatment).

Scaling analysis has primarily relied upon data consisting of either dominance or indifference relations, but not both. It is sometimes the case that both types are collected, but only one analyzed. In one such study by Coombs and Coombs (1977) both types of data were collected and separately analyzed using different methods. It was found that similar scales were obtained under both circumstances. Two implications of this experiment and its results should be mentioned. First, both types of data are meaningful in the sense that they are compatibly generated from the same underlying scale. And second, the measures of quality of the scales

obtained from separate analyses do not reflect the congruence between the subset of the data initially analyzed and the remaining part, ignored in the first analysis, which reflects other aspects of the same scale. In a sense, we are throwing away part of our knowledge, and the measure of quality obtained is therefore inaccurate. In the extreme, the disregarded portion may conflict with the retained part leaving us with faulty conclusions and a meaningless scale.

Graph theorists have handled the problem in a similar manner. In particular, there is a dichotomy in the research dealing with graphs (consisting solely of lines which may represent the indifference relation) and digraphs (consisting solely of directed lines called arcs which have been used to represent the dominance relation). Harary (1969) and Roberts (1979) have thorough treatments of these separate cases, but again no synthesis occurs.

The gist of these arguments is that both types of data should be included in the analysis. We begin with some definitions. Let the stimuli be represented by lower case letters, e.g.,  $x, y, z, w$ , etc. We say that stimulus  $x$  dominates stimulus  $y$ , written  $xDy$ , whenever the number assigned to  $x$  on the unidimensional scale, written  $f(x)$ , is strictly greater than the number assigned to  $y$  by some arbitrary amount, call it  $t$ , the threshold value. In symbols,  $xDy$  if and only if  $f(x) > f(y) + t$ . We say that two stimuli are indifferent, written  $xIy$ , whenever their corresponding numerical values on the unidimensional scale are 'close' to one another, i.e., within the threshold amount  $t$  of each other. In symbols,  $xIy$  if and only if  $|f(x) - f(y)| \leq t$ .

Given these definitions, we wish to determine what characteristics the stimuli need to possess in order to insure a numerical representation on a unidimensional scale. For a finite set of stimuli, the following conditions do the job:

- (1) not  $xDx$
- (2) if  $xDy$  and  $zDw$ , then either  $xDw$  or  $zDy$
- (3) if  $xDy$  and  $yDz$ , then either  $xDw$  or  $wDz$ . (Coombs, Dawes and Tversky, 1970).

This is the representation theorem for semi-orders developed by Scott and Suppes (1958) and first discussed by Luce (1956). These conditions have a simple analog in graphical form, namely that  $xDy$  and  $zIw$  may be represented by an arc from  $x$  to  $y$  and by a line between  $z$  and  $w$ , respectively. The conditions for the representation can be illustrated graphically in the form of restricted or forbidden subgraphs. This set of five forbidden subgraphs may be rewritten in matrix notation, where a one (1) is placed in the  $(i, j)$  cell if stimulus  $i$  dominates stimulus  $j$ , a zero (0) if  $i$  and  $j$  are indifferent, and a minus one ( $-1$ ) if  $j$  dominates  $i$ . In this form it is possible to construct a very simple algorithm to determine whether or not a particular set of data or a specific graph is unidimensional. One such algorithm is as follows. For all stimuli  $i, j, k$ , and  $m$ :

- (1) reject unidimensionality (RU) if any  $(i, i)$  cell contains a one or a minus one
- (2) for every  $(i, j)$  cell containing a one:
  - (a) if  $(j, k)$  contains a one, RU unless  $(i, k)$  also contains a one
  - (b) if  $(j, k)$  and  $(i, k)$  both contain ones, RU if  $(i, m)$ ,  $(j, m)$  and  $(k, m)$  all contain zeros

(c) if  $(k, m)$  contains a one, RU if  $(i, k)$ ,  $(i, m)$ ,  $(j, k)$ , and  $(j, m)$  all contain zeros.

It is hoped that the results presented here will stimulate additional research on topics of mutual concern to psychologists, survey researchers, and mathematicians. One avenue might be to look into other types of data described by Coombs (1976). In fact, Roberts (1979) has already begun. It also seems natural to expand the present results to include some measure of quality, i.e., the minimum number of cells in the matrix which would have to be changed in order to obtain a unidimensional representation, perhaps also taking into account the types of change needed (from a one to a zero seems less severe than to a minus one). The possibilities seem endless.

## References

- C.H. Coombs, *A Theory of Data* (Mathesis Press, Ann Arbor 1976).
- C.H. Coombs and L.C. Coombs, "Don't know": Item ambiguity or respondent uncertainty?, *The Public Opinion Quart.* 41 (1977) 497–514.
- C.H. Coombs, R.M. Dawes and A. Tversky, *Mathematical Psychology* (Prentice-Hall, Englewood Cliffs, 1970).
- F. Harary, *Graph Theory* (Addison-Wesley, Reading, 1969).
- R.D. Luce, Semi-orders and a theory of utility discrimination, *Econometrica* 24, (1956) 178–191.
- F.S. Roberts, Indifference and seriation, *Topics in Graph Theory: Ann. New York Academy of Sciences* 328 (1979) 173–182.
- D. Scott and P. Suppes, Foundational aspects of theories of measurement, *J. Symbolic Logic* 23 (1958) 113–128.